RRCR (2005) 166:89–97

Problems with Using Biomarkers as Surrogate End Points for Cancer: A Cautionary Tale

Arthur Schatzkin

Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd., Suite 320, Room 3040, Rockville, MD 20592-7232, USA
as17h@nih.gov

Abstract Investigations employing surrogate cancer end points are especially attractive because they may be smaller, shorter, and cheaper than comparable studies with explicit cancer outcomes. For many potential surrogate end points—epithelial cell proliferation will be taken as an example—inferences are problematic because of the existence of alternative causal pathways to cancer that bypass the surrogate end point. Evaluating potential surrogates requires information on the following three questions: (1) What is the relation of the surrogate end point to cancer? (2) What is the relation of the intervention (or exposure) to the surrogate? (3) To what extent does the surrogate end point mediate the relation between intervention (exposure) and cancer? Data for these questions may derive from animal experiments, human metabolic studies, observational epidemiologic investigations (including ecologic studies), and randomized trials. Inferences to cancer from such downstream markers as colorectal adenomatous polyps and persistent human papillomavirus infection of the cervix are strong, though not absolutely unassailable. For all but these very-close-to-cancer markers, considerable caution is warranted in extrapolating from surrogate effects or associations to cancer.

Biomarkers can serve three valuable roles in cancer prevention research:

1. They can enhance the biologic plausibility of a particular hypothesis. The recent findings that alcohol intake raises estrogen levels in women [1] provide a plausible physiologic process that could account for the alcohol–breast cancer association observed in many epidemiologic studies [2].
2. Biologic markers of genetic susceptibility may "sharpen" or augment the credibility of a hypothesis.

a. *Sharpening relative risk.* Suppose a given exposure operates only among those with a particular allelic variant of a gene encoding a metabolizing enzyme. Elucidation of the exposure–gene interaction may be critical to observing the relation between the exposure and disease. Stratifying a study population, for example, among those with and without the pertinent metabolizing gene allelotype may reveal a relative risk that would otherwise be obscured in an analysis of the population as a whole.

b. *Enhancing exposure specificity.*

i. *Unraveling mixtures.* Suppose an association exists between a mixture (of foods, industrial agents, etc.) and cancer. If an interaction is observed between the mixture and the gene for a (known) specific metabolizing enzyme, this provides etiologic plausibility for the specific chemical component metabolized by that enzyme.

ii. *Ruling out confounding factors.* For many "lifestyle" exposures the relative risks for cancer are weak to moderate, and confounding is difficult to rule out. Alcohol and breast cancer is an example: the relative risk for moderate alcohol consumption is in the neighborhood of 1.2–1.3 and it is entirely possible that women who drink differ from those who do not in one or more factors that are truly causal for breast cancer. If there were an interaction between alcohol intake and a gene (like ADH3) involved in ethanol metabolism, such that, for example, those who metabolize ethanol more slowly have a qualitatively higher risk of breast cancer compared to more rapid metabolizers, this suggests that it is the alcohol itself, not some confounder, that is a cause of disease (unless the confounding agent is also metabolized by that gene—a not particularly likely scenario).

c. *Mendelian randomization—an antidote for confounding and measurement error* [3]. Certain genetic variants can be viewed as mimicking low or high exposure. In that vein, for example, MTHFR 677TT [4] or HFE [5], respectively, can be viewed as mimicking low folate or high iron exposure. If one of these genetic variants is associated with cancer, this provides independent evidence that the exposure is related to cancer. Because of the random assortment of alleles, this association is not likely due to confounding (though this cannot be ruled out entirely). Moreover, although dietary factors may be assessed with considerable error [6], measurement error is a minimal issue in the genetic association.

3. Finally, biomarkers may serve as surrogates for cancer in epidemiologic studies and clinical trials. This third role is discussed in this article.

1 Why Surrogate End Points?

The occurrence of cancer in humans is a relatively rare event. In the United States, for example, the annual age-adjusted incidence of breast cancer among women is about 100 per 100,000, or 0.1%; the incidence of colorectal cancer among men and women combined is around 50 per 100,000, or 0.05%. Because the incidence of cancer is relatively low, clinical trials and epidemiologic studies must be very large and lengthy (and therefore expensive) in order to accumulate enough cancer cases for meaningful analysis. Studies with surrogate end points are attractive because they can be smaller, faster, and cheaper than those with explicit cancer outcomes.

2 Examples of Potential Surrogate End Points for Cancer

The following are different types of potential surrogate end points for cancer, with specific examples:

- *Tissue*: adenomas; intra-epithelial neoplasia (IEN)
- *Cell*: proliferation, apoptosis
- *Molecular*: DNA adducts, DNA strand breaks
- *Clinical*: imaging end points (e.g., ovarian ultra-sound, mammographic densities)
- *Infection*: human papilloma virus (HPV) infection
- *Blood or urine analytes*: serum or urinary estrogen, prostate-specific antigen (PSA)

3 What Constitutes Surrogate End-Point Validity?

A useful definition of a surrogate end point appears in a preamble to a proposed accelerated approval rule for drugs, from the United States Food and Drug Administration: "A surrogate end point, or 'marker,' is a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions, or survives, and is expected to predict the effect of the therapy" [7]. The essential point, for this article, is that a study of a given intervention (or exposure) in relation to a surrogate end point gives the right answer about the relation of the intervention (exposure) to cancer.

For a surrogate end point to be valid, three conditions must be met:

1. The surrogate is associated with cancer. [The relative risk (RR) is a standard epidemiologic measure of this association.]
2. The exposure is associated with the surrogate. [Relative risk or correlation coefficient (RR, *r*) can be used to reflect this association.]
3. The surrogate mediates the association between exposure and cancer. (The attributable proportion [8] is an indicator of mediation.)

Two examples of the mediation criterion follow [9]:

1. HPV and number of reproductive partners and cervical cancer (Table 1).

Table 1. Number of sexual partners and the risk of cervical dysplasia

	Number of sexual partners				
	1	2	3–5	6–9	>10
Odds ratio					
Unadjusted	1.0	1.7	3.1*	4.7*	4.4*
Adjusted for HPV status	1.0	1.0	1.1	1.5	1.6

HPV, human papillomavirus.
*$p<0.05$.

2. Estradiol and body mass index (BMI) vs breast cancer [10].

Adjusted for free estradiol	RR [95% confidence interval (CI)] for BMI increase of 5 kg/m^2
No	1.19 (1.05–1.34)
Yes	1.02 (0.89–1.17)

Validity is pretty assured for surrogates both necessary for and relatively close, developmentally, to cancer (e.g., CIN3). For other potential surrogates, uncertainty reigns: it is possible to be misled by the existence of alternative pathways to cancer that bypass the surrogate marker (e.g., cell proliferation, as Fig. 1 illustrates).

In the case of hyperproliferation as a potential surrogate end-point biomarker, because an exposure may operate through an alternative pathway (apoptosis) that offsets the pathway through hyperproliferation, hyperproliferation may give the wrong answer about an intervention agent's effect on colorectal cancer in two ways. First, if an intervention reduces (or is associated with lower) proliferation but at the same time reduces apoptosis, then it could have no true effect on colorectal cancer. Second, an intervention

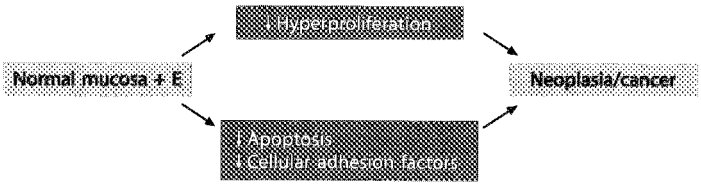

Fig. 1. Alternative pathways from normal colorectal epithelial tissue to neoplasia (adenoma or cancer)

could have no effect on proliferation but could enhance apoptosis, thereby truly lowering the incidence of colorectal cancer. In both these instances, the putative hyperproliferation surrogate marker would give the incorrect answer with respect to colorectal cancer.

Fleming and DeMets showed several examples, from clinical trials comprising intervention, putative surrogate end point, and major clinical end point (e.g., mortality), where the surrogate marker gave a misleading indication of the true clinical end point (mortality, for example) [11].

4 Is Surrogate Marker Validity Generalizable from One Exposure to Another?

A surrogate end point valid for one exposure or intervention vs a cancer is not necessarily valid for a second exposure or intervention. Why? Again, because an alternative pathway to cancer may exist, as Fig. 2 illustrates.

Suppose exposure 1 (*E1*) operates only through the marker *S*. Suppose exposure 2 (*E2*) also operates only through *S*. Then *S* is a valid surrogate for

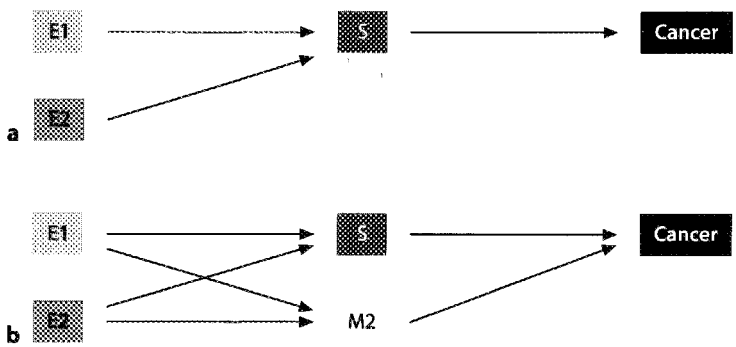

Fig. 2a, b. Surrogate validity for different interventions. In **a** the second intervention (*E2*) operates through the same marker as the first intervention (*E1*); in **b** the second intervention operates through a different marker

both *E1* and *E2*. Now suppose that a second marker (*M2*) exists. If E1 operates primarily through *S*, and only relatively minimally through *M2*, then *S* could still be a valid biomarker for *E1*. But we cannot be certain that *E2* operates similarly through *S* and *M2*—the *M2* pathway may be relatively stronger for *E2*, compared to *E1*, and that *M2* pathway may offset the pathway through *S*.

Thus, we cannot easily be certain that two different intervention agents have pathophysiologic effects so similar that if a given biomarker is a valid cancer surrogate for one agent it must be for the other. That is, we cannot avoid worrying that the second agent has some unanticipated effect on an (unknown?) alternative pathway?

5 Incomplete Validation: The Two-Stage Strategy

In the earlier discussion of validating surrogate end points, three criteria were mentioned: (1) The surrogate is associated with cancer, (2) the exposure is associated with the surrogate, and (3) the surrogate mediates the association between exposure and cancer.

Suppose, for a specific hypothesis, good evidence suggests that criteria (1) and (2) are true. Does this "two-stage" approach ensure that the exposure truly alters (or is associated with) the end point? For example, there are now data indicating that physical activity can lower estradiol levels in women [12]. There are also substantial data now that estradiol levels are directly associated with breast cancer [13]. These two facts, however, do not prove that physical activity necessarily reduces breast cancer risk. A counter-example is instructive: Hormone replacement therapy in women is associated with raised high-density lipoprotein (HDL) levels; HDL is clearly inversely related to cardiovascular disease (CVD) risk. But recent data from the Women's Health Initiative show that hormone replacement therapy (HRT) does not protect against CVD; if anything, HRT use increases CVD risk [14]. Thus, the "two-stage" strategy does not necessarily give the right answer. Why? In this case, it may be that alternative pathways mediating the relation between HRT and CVD offset the pathway through HDL.

6 Colorectal Adenomas as Surrogate End Points for Cancer

Over the last decade, adenoma recurrence trials have become a popular tool for investigating colorectal cancer hypotheses, with interventions ranging from drugs to dietary factors. The rationale for using adenoma recurrence

(defined as the growth of one or more new adenomas after prior detection and removal of one or more earlier adenomas) is:

- Relatively high prevalence in the middle-aged population, so that it is logistically feasible to find and recruit study participants recently diagnosed with a colorectal adenoma (though, in practice, only about 5%–10% of screenees get randomized).
- High recurrence rate (magnitude greater than cancer)—thus, an adenoma recurrence trial can be substantially smaller, faster, and cheaper than a trial with frank cancer end points.
- End-point assessment via standard clinical practice. The investigation can be integrated into standard endoscopic surveillance programs, thereby ensuring reasonably accurate and timely end-point assessment.
- Adenoma–carcinoma sequence. This is the fundamental biologic rationale for adenoma recurrence trials: most colorectal carcinomas develop from adenomas (that are large enough to be detectable at endoscopy).

Nevertheless, adenoma recurrence is not an absolutely conclusive surrogate for colorectal cancer. First, in polyp trials, all participants have had one or more adenomas already. Therefore, there is no information on how the intervention affects early (pre-adenoma) events that could be critical in carcinogenesis. At the other end of the carcinogenesis spectrum, because most recurrent adenomas are small, there is little information provided on later events, either (1) growth of small, non-advanced to large or advanced adenomas or (2) transformation of large or advanced adenomas to carcinoma. The intervention could have a critical impact on these early or late events, but the adenoma recurrence end point in the trial will not permit this impact to be evaluated.

Second, adenomas may be heterogeneous lesions, such that only a small subset eventually go on to cancer. An intervention can have different effects on the "innocent" and "bad" adenomas. If the intervention affects only innocent adenomas, because such lesions constitute the very large majority of recurrent adenomas, it would appear that the intervention lowers adenoma recurrence and, by inference, would reduce colorectal cancer incidence, when in fact the intervention has no effect on cancer. Alternatively, suppose an intervention affects only bad adenomas, not the innocent ones (Fig. 3).

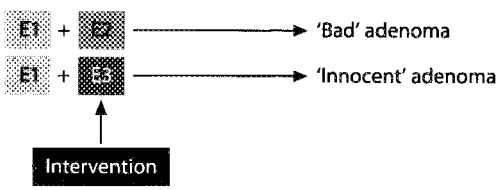

Fig. 3. Limitation of colorectal adenoma recurrence as a surrogate for colorectal cancer

Then it would appear that the intervention has only a minimal impact on adenoma recurrence (and, therefore, cancer), when the intervention really does reduce cancer incidence.

Alternative approaches to adenoma recurrence trials have been proposed, e.g., only persons with a prior large or advanced adenoma could be eligible for the trial. The rationale is that such participants are more likely than those with only a small, non-advanced lesion to have some sort of pre-malignant "field defect." There are serious cost implications to this strategy, however. Because only about one-third of participants in polyp trials have had a large/advanced lesion, three times as many individuals have to be screened to achieve the target sample size for the study. A second alternative strategy is to have as the end point only advanced adenomas (those 1+ cm, or with villous elements or high-grade dysplasia). The rationale is that the heterogeneity of recurrent adenomas is reduced: a much larger proportion of such end-point lesions is likely to progress to invasive cancer. But, because only some 1/6 of recurrent lesions in standard polyp trials are advanced, the sample size of the trials will have to be increased substantially—leading to a much more expensive investigation.

In summary, results of adenoma recurrence trials constitute strong—but not absolute—evidence on colorectal cancer. In addition, such trials are not inexpensive. Costs rise with alternative designs intended to strengthen inference.

7
Statistical Considerations

There is substantial variability ("noise") in biomarker measurements, with several sources of within-participant variation (e.g., over time, between specimen collections, reading-to-reading). The "signal-to-noise" ratio may be problematic: If within-person variation is large, it may not be possible to discriminate among participants. It may be possible to decrease within-participant variation by taking repeat samples (e.g., more biopsies, multiple blood samples). Information on variance components is critical. Such data are sparse (estradiol, proliferation). Measurement error will attenuate associations between exposure and marker, on the one hand, and marker vs cancer, on the other. Moreover, measurement error can lead to underestimation of extent to which surrogate mediates the effect of exposure on cancer

8 Conclusions

There are at least two ironies accompanying the surrogate end-point problem. First, the large, long, costly studies needed for evaluation are precisely the studies surrogates were designed to replace. Second, inferential certainty is directly associated with study cost—that is, you get what you pay for. Nevertheless, surrogate end points may be particularly valuable in phase II studies. And, in conjunction with other types of investigations (e.g., polyp trials plus cohort studies of colorectal cancer), such surrogate end-point markers may markedly enhance the "probability of being right" about the etiology and prevention of cancer.

References

1. Reichman ME, Judd JT, Longcope C, Schatzkin A, Nair PP, Campbell WS, Clevidence BA, Taylor PR (1993) Effects of moderate alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. J Natl Cancer Inst 85:722–727
2. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW Jr, et al (2002) Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 87:1234–1245
3. Davey-Smith G, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32:1–22
4. Little J, Sharp L, Duthie S, Narayanan S (2003) Colon cancer and genetic variation in folate metabolism: the clinical bottom line. J Nutr 133:3758S–3766S
5. Shaheen NJ, Silverman LM, Keku T, Lawrence LB, Rohlfs EM, Martin CF, Galanko J, Sandler RS (2003) Association between hemochromatosis (HFE) gene mutation carrier status and the risk of colon cancer. J Natl Cancer Inst 95:154–159
6. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano R, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ (2003) The structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol 158:14–21
7. New drug, antibiotic and biological drug product regulations: accelerated approval (1992) Proposed Rule. 57: Federal Register 13234–13232
8. Rothman KJ, Greenland S (1998) Modern epidemiology. Lippincott-Raven, Philadelphia
9. Schiffman MH, Schatzkin A (1994) Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 54:S1944–S1947
10. Key TJ, Appleby PN, Reeaves GK, Roddam A, Dorgan JF, Longcope C, et al (2003) Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. J Natl Cancer Inst 95:1218–1226
11. Fleming T, DeMets DL (1996) Surrogate end points in clinical trials: are we being misled? Ann Intern Med 125:605–613
12. McTiernan A, Kooperberg C, White E, Wilcox S, Coates R, Adams-Campbell LL, Woods N, Ockene J (2003) Recreational physical activity and the risk of breast cancer

in postmenopausal women: the Women's Health Initiative Cohort Study. JAMA 290:1331–1336
13. Key TJ (1999) Serum oestradiol and breast cancer risk. Endocr Relat Cancer 6:175–180
14. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321–333